AERES report on the research unit

Modélisation Conceptuelle des Connaissances Biomédicales

From the

Université de Rennes 1

INSERM

December 2010

Section des Unités de recherche

AERES report on the research unit

Modélisation Conceptuelle des Connaissances

Biomédicales

From the

Université de Rennes 1

INSERM

Le Président de l'AERES

Didier Houssin

Section des unités
de recherche

Le Directeur

Pierre Glorieux

December 2010

Research Unit

Name of the research unit: Modélisation Conceptuelle des Connaissances Biomédicales

Requested label: UMR S INSERM

Name of the director: Ms. Anita BURGUN

Members of the review committee

Committee chairman

M. Jacques DEMONGEOT, Université Joseph Fourier, Grenoble

Other committee members

Ms. Catherine DENEUX, Université Pierre et Marie Curie, Paris

Ms. Chantal REYNAUD, Université Paris Sud, Orsay

Observers

AERES scientific advisor

M. Pierre LEGRAIN

University and Research Organization representatives

Ms. Marie-José Leroy, INSERM

Report

1 • Introduction

• Date and execution of the visit

The visit took place on December 13, 2010 at the Laboratory U 936 INSERM "Conceptual modelling of biomedical knowledge". After an introduction by the Director, A. BURGUN and a closed meeting between the Committee members, the review of the past activity during the two last years, and then the future scientific project of the unit have been presented successively by A. BURGUN and O. DAMERON. After this general presentation, three actions have been detailed:

- Mining electronic health Records for improving CT recruitment: Web Semantic approach
- Computer assisted decision in tele-cardiology
- Inter-specific comparison between metabolic pathways.

The visit ended with meetings planned by the AERES regulation, with researchers, post-docs, PhD students, the ITA and IATOS personnel, and eventually with the representatives of the University and of INSERM.

• History and geographical localization of the research unit, and brief presentation of its field and scientific activities

This unit has been created in 2009 and is located at the Faculty of Medicine of Rennes inside of the Health Campus of Pontchaillou. The main research topic of the unit is the conceptual modelling of the biomedical knowledge in the framework of dedicated ontologies, necessary step before building Data and Knowledge Base Management Systems, or Data and Knowledge Warehouses and Repositories.

• Management team

The management team includes:

- a director : A. BURGUN,

- a deputy director : O. DAMERON

- a secretary manager : D. GUILLOTIN

• Staff members (on the basis of the application file submitted to the AERES)

	Past	Future
N1: Number of researchers with teaching duties (Form 2.1 of the application file)	6	5
N2: Number of full time researchers from research organizations (Form 2.3 of the application file)	0	0
N3: Number of other researchers including postdoctoral fellows (Form 2.2 and 2.4 of the application file)	3	4
N4: Number of engineers, technicians and administrative staff with a tenured position (Form 2.5 of the application file)	0.6	0.6
N5: Number engineers, technicians and administrative staff without a tenured position (Form 2.6 of the application file)	3.5	
N6: Number of Ph.D. students (Form 2.7 of the application file)	6	
N7: Number of staff members with a HDR or a similar grade	3	2

A meeting was held with some of the researchers of the group, which consists of permanent Faculty of Medicine and University Hospital members corresponding to 3 professors, 3 assistant professors, and 3 University Hospital assistants. There are no full-time research scientists or research directors. The group makes significant contributions to teaching. Two of the members are responsible of a Master2. Three of the members have their Habilitation à Diriger des Recherches (HDR). One of them will retire soon. Three assistant professors of the group should receive their HDR in the next 4 years.

Researchers suffer from a recruitment problem concerning both permanent research members and engineers. New permanent positions have not been created since 2007. It is essential to rapidly strengthen the team with new University or INSERM members to assist both teaching and research activities, to ensure the viability of the group and the success of the many projects it is involved in. Recruitment difficulties are due to the lack of positions but also to the lack of attractiveness of the team. So, the researchers would strongly support a restructuring of their group.

A joined operation with other health STIC research groups would be for th researchers an opportunity to increase the size of the team and consequently its visibility and attractiveness. Collaborations have been developed in the last two years with some of these teams located in Rennes (Unité VisAGeS and LTSI) and a restructuration implying these teams seems to constitute a logical solution to the visibility and attractiveness issues.

The meeting with ITA and IATOS personnel has shown a great motivation and involvement in the life and research activities of the unit, but most engineers come from the University Hospital and not from the Faculty of Medicine. Furthermore, one engineer will leave in 2011 and the position will be lost. The rapid recruitment of a research engineer coming from INSERM of from the University is critical.

2 • Overall appreciation on the research unit

- ### Summary

Despite the weak number of permanent researchers (6) and the absence of full-time researcher without teaching and/or hospital duties, the scientific productivity is very good in specialized journals of the domain (medical informatics). The unit represents the only French centre having an international visibility on the structuration of medical terminology and biomedical knowledge in the framework of ontologies, necessary before any data or knowledge organized repository (like data or knowledge bases and warehouses) building in medicine.

- ### Strenghts and opportunities

The monopolistic situation of the unit is very favourable to be recognized as the national reference center in biomedical ontologies, whose expertise could be used by French Faculties of Medicine, University Hospitals and research institutions like INSERM and CNRS. The presence of the unit in several FP7 European projects is a good opportunity to enlarge its international visibility.

- ### Weaknesses and threats

The unit needs a rapid recruitment in terms of full-time researcher (e.g., at INSERM or CNRS), as well as the consolidation (in terms of academic position) of at least one non-permanent research engineer. If not, the enthusiasm and energy of the start could disappear causing dangerous damages to the present creativity and solidarity of the unit. The papers corresponding to conferences or oral presentations in excellent international conferences (AMIA, Medinfo,…) could be often doubled by publications in the best international journals of the domain of the unit, medical informatics.

- **Recommendations**

The research axes of the unit are very promising and have to be deepened and reinforced so that the unit could become the reference in terms of medical knowledge organization and management in Europe. A reflexion about the possible convergence with other research units on the Rennes Health site must be pursued during the next quinquennial period.

- **Production results**

The scientific production is very good by taking into account the weak number of permanent researchers and the absence of full-time researchers in the unit.

A1: Number of permanent researchers with teaching duties (recorded in N1) who are active in research	6
A2: Number of permanent researchers without teaching duties (recorded in N2) who are active in research	0
A3: Ratio of members who are active in research among staff members [(A1 + A2)/(N1 + N2)]	1
A4: Number of HDR granted during the past 4 years	0
A5: Number of PhD granted during the past 4 years	8

21 scientific papers produced by the unit during the 5 past years (2006-2010) have been published essentially in very good journals of the disciplines "medical informatics", "bio-informatics" and "medical imaging", plus 22 papers from excellent international conferences (AMIA, Medinfo,…), plus some articles on the clinical side, which corresponds to a very good research activity for only 6 permanent researchers with teaching duties:

- 3 in BMC Bio-informatics (ISI IF 3.43 ; ISI mathematical and computational biology journal ranking 4/29)

- 4 in Int. J. in Medical Informatics (ISI IF 3.13 ; ISI medical informatics journal ranking 3/23)

- 2 in J. of Biomedical Informatics (ISI IF 2.43 ; ISI medical informatics journal ranking 6/23)

- 3 in Artificial Intelligence in Medicine (ISI IF 1.65 ; ISI medical informatics journal ranking 13/23)

- 1 in Medical Informatics Internet Med. (ISI IF 1.04 ; ISI medical informatics journal ranking 17/23)

- 1 in J. Nuclear Medicine (ISI IF 6.42 ; ISI radiology, nuclear medicine and medical imaging journal ranking 1/104)

- 1 in Eur. J. Nuclear Med. (ISI IF 4.5 ; ISI radiology, nuclear medicine and medical imaging journal ranking 7/104)

- 1 in Am. J. of Neuroradiology (ISI IF 3.3 ; ISI radiology, nuclear medicine and medical imaging journal ranking 16/104)

- 1 in Radiation Oncology (ISI IF 2.53 ; ISI radiology, nuclear medicine and medical imaging journal ranking 32/104)

- 4 in J. of Radiology (ISI IF 0.63 ; ISI radiology, nuclear medicine and medical imaging journal ranking 94/104)

- 14 in Studies in Health Technologies and Information

- 5 in AMIA Annual Sympsium Proceedings

- 2 in Lecture Notes in Computer Science

The director of the unit has a good international visibility in its research field, i.e., medical informatics: 8 as ISI h-index and 19 as Google Scholar h-index (the best ISI h-index in this field in France is equal to 22).

3 • Specific comments

• Appreciation on the results

The research done in this unit in the domain of biomedical knowledge management is original, generic and promising. It is already used in several projects of medical data and knowledge bases or warehouses. The quality of the publication activity is attested by not only the number but also the notoriety of the journals and congresses in which appear the papers written by the researchers of the unit. The impact in terms of patents and exploitation licensing is weak, but an interesting industrial partnership is present in the European FP7 project EHR4CR with pharmaceutical industry, as well as in the project Akenaton and its extension "Off-site Cardiology" with the European company SORIN and could occur from the project Astec with the regional Cancero-Pole Onco-Bretagne and from the project "Interspecies comparison of metabolic pathways" with the national institution INRA devoted to agronomic research. Ancient collaborations exist about the tools for combining voice recognition and automatic indexing of medical reports with the company Intermède and a new partnership is starting with the company IBC (Integrative BioComputing) for the development of serious games in the framework of the French Virtual Medical University.

• Appreciation on the impact, the attractiveness of the research unit and of the quality of its links with international, national and local partners

The unit has an international visibility allowing welcoming PhD students and post-doctoral fellows in the framework of FP7 European projects. The ancient and deep link with the National Library of Medicine ensures also a good flux of international collaborations and invitations in excellent international congresses in medical informatics like AMIA and Medinfo.

• Appreciation on the management and life of the research unit

The unit is well managed and the stabilization of a secretary devoted to helping the director in administrative tasks has been a good decision. Three promising cutting edge projects have emerged during the two past years, "Mining electronic health Records for improving CT recruitment: Web Semantic approaches", "Off-site Cardiology" and "Interspecies comparison of metabolic pathways". The staff members are involved in teaching in 2 research masters (MTIB-H & Bioinfo) and participating to the scientific life of the local of the Doctoral School VAS (Vie-Agro-Santé).

• Appreciation on the scientific strategy and the project

The close link between the hospital information system and the health research area favored by the involvement of INSERM is a strong point of the group. The research deals with very important, difficult and challenging problems in knowledge methodology that arise in many applications including biomedical ones when we have to integrate heterogeneous data. These problems have a great interest both from a scientific point of view and in terms of real-life biomedical applications. The solutions involve construction of models or ontologies and their use for semantic interoperability. The team presents significant contributions in this field.

The scientific contributions of the unit U 936 are guided by a constant strategy taking into account application requirements, searching systematically to detect if they are not specific of a given application in order to reach a global and high level genericity allowing their reuse (with a minor reengineering) in other projects related to various biomedical themes. The team has acquired for many years a solid expertise in the UMLS. Also it is of a small size, its dynamism is well established. Active and well-chosen collaborations at international level have been developed with NIH/NLM, Stanford and Edinburgh Universities.

The unit is involved in several funded research projects at national level (through ANR, ACI or regional funding) and at EU level (through 2 FP7 projects recently obtained). The number of different projects is impressive. Ontologies and software systems have been produced and made available on the Web.

The team has a clear scientific strategy. Its project is in the continuity of the research work being currently carried out. The aim is to consolidate and to develop it. One direction is to explore the applicability of its results in the context of data warehouse operational applications and to ensure a faster exchange of information flows between research and care units.

The project is articulated around two main issues: integration of biomedical data and improvement of reasoning mechanisms on these data. Both issues will often be studied in the same research project related to a given medical topic. On the integration issue, the problems of establishing relationships between data elements and ontologies, multi-domain and multi-scale knowledge integration and similarity measures will be investigated. All these problems are interesting and very relevant. The team possesses the necessary knowledge to achieve excellent results. On the reasoning issue, the major focuses will be on information at different levels of granularity, cooperation of rules and DL in an optimal way, exploiting non formal relationships and medical context. All these problems will be studied in the setting of biomedical applications but they arise in many other fields. They are important and challenging issues. The team has the skills to contribute to innovative solutions on these aspects too.

Integration and reasoning problems will be investigated through three main applications: off-site cardiology, mining electronic health records for improving recruitment in clinical tests, and interspecies comparison of metabolic pathways. These health problems are very important because they provide real requirements and allow evaluation of solutions by real users. Dealing with real problems is very often complex. The team manages very well that point thanks to the association of researchers in computer science, knowledge management, ontology engineering, semantic web, who are also experts in the health domain.

To conclude, the relevance and feasibility of a long term (5 years) scientific project is clearly proved, the resources (in particular those coming from external sources as ANR and EU) are ensured and the 3 specific projects presented are sufficiently original and necessary to guarantee a place quite apart for the unit in the French community of medical informatics.

4 ● Appreciation project by project

1) Mining electronic health Records for improving CT recruitment : Web Semantic approaches

The project concerns the use of ontology to represent both patient data and eligibility criteria for entering in cohorts for multi-site health inquiries. The project implies the development and integration of different tools:

- a unique terminology identifier for concepts

- tools of subsumption reasoning in a proper ontology

- SWRL (Semantic Web Rule Language) to be combined with OWL

- a reasoning tool over data types and time intervals

- use of NCI Thesaurus (National Cancer Institute USA)

- good coverage of domain specific terminologies

- development in Open Source (no licence fees like for SNOMED)

The first developments have been already converted in OWL (Ontology Web Language) Land the research activity is pursued in the same direction. For example, the knowledge embedded in the ontology allowing for example to recognise the prostate adeno-carcinoma as a malignant neoplasm.

This project is promising and should improve the relationship already existing between the unit and the clinicians of the University Hospital if it develops new end-user interface tools. The knowledge representation requires indeed effort: needing to stay within the boundaries of SWRL expressivity and idioms, implying the use of: i) no explicit disjunction, needing to use separate rules, ii) a match of an instance in a relation, and then iii) type checking methods. The criteria are written in a higher level representation, and then automatically compiled into SWRL.

The project is important and follows the successful project Astec of reuse of patient data in pharmaco-genetic and pharmaco-physiologic cohorts.

2) Off-site Cardiology

This project is the following of the project AKENATON (Automated Knowledge Extraction from medical records iN Association with a Telecardiology Observation Network). After a proof of concept, the project aims in collaboration with the company SORIN (implanted cardiac devices) to achieve:

- development of an ontology in cardiology including medical reasoning rules represented in SWRL

- extraction of clinical information from patient data files and exit letters

- contextualisation of alarms from clinical data coming from portable devices (like permanent pacemakers)

- integration of clinical data within existing standards and creation of norms from the French register of cardiac pathologies of rhythm (SFC)

- at home surveillance of pacemaker implanted patients

- materio-vigilance long-term testing

This project would lead to a European standard co-developed with the major European company of the domain (SORIN) for following and watching at home cardiac patients with risks in major troubles of the cardiac rhythm (tachy- and bradycardiopathies, fibrillation,…).

3) Interspecies comparison of metabolic pathways

This new project concerns a comparative study of the control of the lipid metabolism in collaboration with INRA. It involves a work on animal models to extract a common explanatory knowledge about the lipid regulation, and then start an application on human implying therapeutic proposals.

The main objectives of the project are:

- To develop generic methods for comparing metabolic pathways between species on both qualitative and quantitative sides, using the Wang semantic distance (after Wang et al. in Bioinformatics 2007)

- Validate these methods exploiting new experimental data (chicken), existing annotated data (human and mice) and knowledge bases (like BioCyc for biochemical reactions, GDB or UniProt, in which a gene or an enzyme "annotated to a term can be retrieved not only with this term, but also with all of its parent terms, increasing flexibility and power when searching for and making inferences about genes" as noticed by Rhee et al. in Nature Genetics 2008), ontologies (like GO, GeneOntology, for organizing genomic concepts) and specialized literature.

This project will offer a new tool of information query on the web, called Go2Pub, already more performing than Pubmed in some contexts. The approach proposed is original and could be capitalized in the lipid context and, then, repeated and exploited in other metabolic contexts.

Intitulé UR / équipe	C1	C2	C3	C4	Note globale
MODÉLISATION CONCEPTUELLE DES CONNAISSANCES BIOMÉDICALES	A+	A+	A+	A+	A+

C1 Qualité scientifique et production

C2 Rayonnement et attractivité, intégration dans l'environnement

C3 Gouvernance et vie du laboratoire

C4 Stratégie et projet scientifique

Statistiques de notes globales par domaines scientifiques
(État au 06/05/2011)

Sciences du Vivant et Environnement

Note globale	SVE1_LS1_LS2	SVE1_LS3	SVE1_LS4	SVE1_LS5	SVE1_LS6	SVE1_LS7	SVE2 _LS3 *	SVE2_LS8 *	SVE2_LS9 *	Total
A+	7	3	1	4	7	6		2		30
A	27	1	13	20	21	26	2	12	23	145
B	6	1	6	2	8	23	3	3	6	58
C	1					4				5
Non noté	1									1
Total	**42**	**5**	**20**	**26**	**36**	**59**	**5**	**17**	**29**	**239**
A+	16,7%	60,0%	5,0%	15,4%	19,4%	10,2%		11,8%		12,6%
A	64,3%	20,0%	65,0%	76,9%	58,3%	44,1%	40,0%	70,6%	79,3%	60,7%
B	14,3%	20,0%	30,0%	7,7%	22,2%	39,0%	60,0%	17,6%	20,7%	24,3%
C	2,4%					6,8%				2,1%
Non noté	2,4%									0,4%
Total	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%

* les résultats SVE2 ne sont pas définitifs au 06/05/2011.

Intitulés des domaines scientifiques

Sciences du Vivant et Environnement

- **SVE1 Biologie, santé**
  - **SVE1_LS1 Biologie moléculaire, Biologie structurale, Biochimie**
  - **SVE1_LS2 Génétique, Génomique, Bioinformatique, Biologie des systèmes**
  - **SVE1_LS3 Biologie cellulaire, Biologie du développement animal**
  - **SVE1_LS4 Physiologie, Physiopathologie, Endocrinologie**
  - **SVE1_LS5 Neurosciences**
  - **SVE1_LS6 Immunologie, Infectiologie**
  - **SVE1_LS7 Recherche clinique, Santé publique**
- **SVE2 Ecologie, environnement**
  - **SVE2_LS8 Evolution, Ecologie, Biologie de l'environnement**
  - **SVE2_LS9 Sciences et technologies du vivant, Biotechnologie**
  - **SVE2_LS3 Biologie cellulaire, Biologie du développement végétal**

PRÉSIDENCE

Guy Cathelineau
Président

Rennes, le 12 avril 2011

Monsieur Pierre GLORIEUX
Directeur de la section des unités de recherche
Agence d'Evaluation de la recherche et de
l'Enseignement Supérieur (AERES)
20, rue Vivienne
75002 PARIS

Vos réf. : S2UR120001340
U936 MCCB– 0350936C

Monsieur le Directeur,

Je vous adresse mes remerciements pour la qualité du rapport d'évaluation fourni à l'issue de la visite du comité d'expertise concernant l'unité mixte de recherche «**Modélisation Conceptuelle des Connaissances Biomédicales** ».

L'université de Rennes 1 sera particulièrement attentive à ce que les recommandations formulées par le comité de visite soient prises en compte.

A la lecture de ce rapport, vous trouverez ci-joint, les réponses du directeur d'unité auxquelles nous souscrivons en totalité, en y ajoutant quelques précisions sur les deux éléments suivants :

La thématique aux interfaces STIC-Santé est un des axes stratégiques et intersectoriels fortement développé au sein de l'Université de Rennes 1, en partenariat avec l'INSERM (trois unités communes) et le CHU. La perspective d'actions fédératrices, visant à mutualiser et à rendre encore plus visible ce potentiel de recherche, est à renforcer tout au long du prochain contrat quinquennal.

Dans ce contexte favorable, la jeune unité U936 Inserm créée en 2009, a pris toute sa place et un rôle affirmé en information médicale, ontologie et systèmes à base de connaissances. Son potentiel scientifique se doit d'être renforcé dans les années futures par des recrutements de chercheurs permanents.

Je vous prie d'agréer, Monsieur le Directeur, l'expression de ma considération distinguée.

Le Président de l'Université de Rennes 1

Guy CATHELINEAU

P.S : Nous souhaitons que soit ajouté dans le rapport la mention de la présence de Monsieur Claude LABIT (Université de Rennes 1) dans la liste des « University and Research Organization Representatives »

2, RUE DU THABOR
CS 46510
35065 RENNES CEDEX
www.univ-rennes1.fr

TÉL. 33 (0)2 23 23 36 36
FAX 33 (0)2 23 23 36 00

president@univ-rennes1.fr

U 936

Modélisation des connaissances biomédicales
Professeur Anita Burgun, Directeur

Objet : Rapport d'Evaluation de l'Unité de Recherche U936,
Modélisation des Connaissances Biomédicales

Madame, Monsieur,

L'opinion du directeur de l'U936 favorable à un regroupement avec les autres unités STIC Santé du site est partagée par les membres de l'unité. Les discussions en ce sens amorcées en 2010 avec les directeurs et les responsables d'équipes de VISAGES et du LTSI se poursuivront.

Nous sommes d'accord sur le fait que plus de publications dans des journaux auraient pu être soumises. Au cours des trois derniers mois, trois nouveaux articles ont été acceptés –deux dans Methods of Information in Medicine (IF 1.69), et un dans IJMI (IF 3.13), et au moins deux articles sur l'ANR TecSan AKENATON (coord. A. Burgun, projet se terminant le 31 mai) vont être soumises prochainement à des revues. De plus, nous pensons que notre participation à 2 projets européens FP7--TRANSFoRm (participation débutée en Sept 2010) et EHR4CR (début en Jan 2011) et la création du labex COMIN Labs porté par le PRES Université Européenne de Bretagne, dont un des axes est l'exploitation des Technologies de l'Information et de la Communication (TIC) pour une médecine améliorée et personnalisée, offriront de nouvelles opportunités de recherche, de valorisation et de publication.

Fait à Rennes le 11 avril 2011

Professeur Anita BURGUN